Optimal Transport for Secure Spread-Spectrum Watermarking of Still Images

Benjamin Mathon, François Cayre, Patrick Bas, Benoît Macq

Optimal Transport for Secure Spread-Spectrum Watermarking of Still Images

Benjamin Mathon, François Cayre, Patrick Bas, *Member, IEEE* and Benoît Macq, *Fellow, IEEE*

Abstract

This article studies the impact of secure watermark embedding in digital images by proposing a practical implementation of secure spread-spectrum watermarking using distortion optimization. Because strong security properties (key-security and subspace-security) can be achieved using Natural Watermarking (NW) since this particular embedding lets the distribution of the host and watermarked signals unchanged, we use elements of transportation theory to minimize the global distortion (MSE). Next, we apply this new modulation, called Transportation Natural Watermarking (TNW), to design a secure watermarking scheme for grayscale images. TNW uses a multiresolution image decomposition combined with a multiplicative embedding which is taken into account at the distribution level. We show that the distortion solely relies on the variance of the wavelet subbands used during the embedding. In order to maximize a target robustness after JPEG compression, we select different combinations of subbands offering the lowest BERs for a target PSNR ranging from 35 to 55 dB and we propose an algorithm to select them. The use of transportation theory also provides an average PSNR gain of 3.6 dB on PSNR with respect to the previous embedding for a set of $2,000$ images.

This work was supported in part by the National French project ANR-10-CORD-019 Estampille, the Belgian project BCRYPT IAP (PAI) phase-VI and by the installation grant of Prof. S. Voloshynovskiy, University of Geneva, Switzerland.

B. Mathon is with GIPSA-LAB, Grenoble-INP, 38402 Saint Martin d'Hères cedex, France, and also with CVML, SIP group, University of Geneva, 1227 Carouge 4, Switzerland (email: benjamin.mathon@grenoble-inp.fr).

F. Cayre is with GIPSA-LAB, Grenoble-INP, 38402 Saint Martin d'Hères cedex, France (email: francois.cayre@gipsa-lab.grenoble-inp.fr).

P. Bas is with CNRS LAGIS, École centrale de Lille, Avenue Paul Langevin BP48, 59651 Villeneuve d'Ascq cedex, France (email: patrick.bas@ec-lille.fr).

B. Macq is with TELE, Université catholique de Louvain, Bâtiment Stévin 2, Place du Levant, 1348 Louvain-la-Neuve, Belgium (email: benoit.macq@uclouvain.be).

Index Terms

Watermarking, Digital images, Transportation theory

I. Introduction

The development of multi-bit watermarking techniques, which consist in embedding a message (several digits, often binary symbols) in a host content such as a movie, a song, a text or a picture, has followed different paths of development since the pioneering studies. Initially, the main goal for the community was to design a watermarking scheme which can resist common signal processing operations. The embedded mark had to be detected after geometrical transformations, compressions, Gaussian noise addition, *etc*. The second intuitive constraint was that the presence of the watermark should remain imperceptible. These two constraints, respectively *robustness* and *imperceptibility*, are very important for a watermarking scheme but the major problem for researchers is the non-orthogonality of these constraints. In fact, in order to be robust, the mark has to be embedded with a strong power but this power implies a strong distortion of the host content. Combining robustness and imperceptibility has been a challenge for several years [1]–[4], which motivated the development of watermarking based on spread-spectrum modulations [5], [6] or quantization techniques [7]–[9].

A constraint receiving more and more attention in watermarking nowadays is the *security* constraint, defined as "the inability by unauthorized users to have access to the raw watermarking channel" by Kalker [10] and linked with the presence of an adversary [11]–[16]. Watermarking schemes then are required to respect the Kerckhoffs' principle [17] (*i.e.* a secret key is shared between the encoder and the decoder and is the only secret parameter for the adversary [18], [19]). The Kerckhoffs' principle comes from cryptography and cryptanalysis but is also widely used in watermarking [20], [21] or in other forms of data-hiding like steganography [22]–[24]. However, watermarking security analysis shows that an adversary may estimate the secret if he owns a sufficient number of marked contents and if the embedding scheme is not secure [11]–[16]. According to the degree of estimation of the secret, the adversary would be able to erase, modify or copy the embedded message to another host content. There are two major differences between robustness and security. On one side, robustness attacks are deemed not intentional: signal processing operations on watermarked contents like compression can be done by the provider before the legal use by a user of this content. On the other side, security attacks come from an adversary who deliberately wants to hack the watermarking scheme, such attacks are deemed more harmful than robustness attacks because the adversary can both alter the embedded message and perform

an optimal minimization of the attack distortion [25] or even recover the original image [26].

In order to understand the ins and outs of the security game in watermarking, it is important to detail the role of the three important actors in the data hiding process:

1) the *distributor* is the person who marks the host contents. He is the only one who knows the secret key used for embedding and decoding. The distributor can be the author of the content, the company who sells the content or a trusted third party. His main important constraints are robustness and security,
2) the *user* is a person with privileges related to the content, and set by the distributor. For examples: reading a digital book, listening to a song or watching a video. His most important constraint is the imperceptibility of the watermark,
3) the *adversary* will try to estimate the secret key in order to modify the watermark. Because of this modification, he can obtain more privileges than a simple user (unauthorized by the distributor). For example, in a fingerprinting scenario: he can resale the content in an illegal way, share the content on peer-to-peer networks, *etc*. The adversary can be a (group of) user(s) or corrupted user(s).

Robustness, imperceptibility and security are the three constraints that a distributor has to take into account for the development of a watermarking scheme. More precisely, the distributor looks for an optimization of one constraint while settling the others. Moreover, the importance of each constraint depends on the use case. The trade-off between imperceptibility, robustness and security is represented by the triangle of constraints in watermarking on Fig. 1.

In this work, we are interested in optimizing robustness and imperceptibility constraints when security is set first. The security of a watermarking scheme depends on the assumptions related to the use case under consideration. Security attacks on watermarking are mainly linked with information leakage: what are the data that the adversary can have access to?

We consider here the Watermarked Only Attack framework (WOA) [21], which means that an adversary has access to both the source code of the scheme (without knowing the secret key) and $N_o \geq 1$ contents watermarked with the same, unknown key. Gathering information given by the marked contents, his goal will be to estimate the secret key, *i.e.* a set of secret watermark components, that were initially used during embedding. If this is not possible, the adversary can try to estimate a less informative secret parameter such as, for example, the private subspace spanned by these components in the case of spread-spectrum embedding.

It is important to point out that security in WOA depends on the distribution of contents after watermarking. We recall the three main classes of security in WOA [21] by increasing security level:

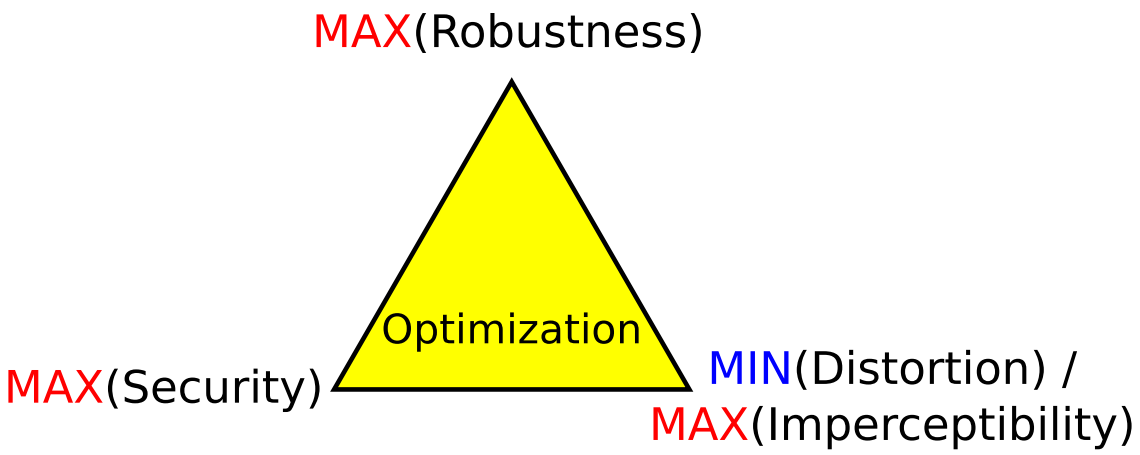


Figure 1. The triangle of constraints in watermarking where the distributor looks for an optimization of one constraint while settling the two others. The importance of each constraint depends on the use case scenario and the distributor has to take into account the trade-off between the three constraints.

- *Insecurity*: the distribution of marked contents is unique and specific to the chosen secret key. More precisely, two different keys generate two different distributions of marked contents. An adversary can then estimate the secret components [27].
- *Key-security*: for a subset of keys, the distribution of the marked contents will be the same. With a sufficient number of observations, the adversary will be able to estimate a subset of the watermarking space (a subspace in the case of spread-spectrum watermarking).
- *Subspace-security*: the distribution of the marked contents is the same for each possible key. The adversary cannot estimate a set smaller than the original set of possible watermark signals.

Spread-spectrum modulations such as Natural Watermarking (NW) [28] and Circular Watermarking (CW) [29] have been designed to deter a correct estimation of the secret components: NW prevents the whole subspace estimation and is consequently subspace-secure, while CW prevents the estimation of the watermark components and therefore qualifies for key-security.

In this article, we propose a practical implementation of NW for still images. We look for an optimization of the distortion constraint that reduces the distortion due to the assignment of each host content to a marked content. This problem finds a solution using transportation theory [30], a research domain that was successfully used for the development of watermarking techniques as in [31] (Transportation NW modulation), [32] (Soft-scalar Costa Scheme) and [33] (Controllable Secure Watermarking). We implement the Transportation NW modulation (TNW) on multiple combinations of wavelet subbands of grayscale images and we add a perceptual weighting in order to increase the imperceptibility of the

whole system. We show that our optimization of imperceptibility can be made while keeping the desired security level. This work is a practical extension of [31], where the transportation problem was presented and applied only to synthetic Gaussian signals, and we focus here on the design of secure embedding on still images.

In Sec. II, we illustrate the links between subspace-secure watermarking and the distributions of marked contents using two embedding techniques. We present elements of transportation theory which optimize assignments between host and marked distributions in a theoretical way, and we further use these results to design TNW. In Sec. III, we tackle implementation issues and propose an experimental watermarking scheme in the wavelet domain with psycho-visual masking. Since the robustness and the distortion of the proposed scheme are intimately linked with the set of the selected subbands, we present the combinations which provide the best robustness for a given distortion, computed from $2,000$ images.

II. Distribution matching watermarking

We first list the notations and conventions used in this article. Functions are denoted in roman fonts, sets in calligraphy fonts, vectors and matrices in bold fonts and variables in italic fonts. Vectors are written in small letters and matrices in capital ones. The content of a vector $\mathbf{x}$ with length n is denoted by $(\mathbf{x}(0)\dots\mathbf{x}(n-1))$, the random variable associated to i.i.d components is denoted X. P_δ (resp. f_δ) is the cumulative distribution function (resp. probability density function) of a distribution δ. $|.|$ is the absolute value on $\mathbb{R}$ and $\|.\|$ is the Euclidean norm on $\mathbb{R}^N$ with $N \geq 1$. $\langle .|. \rangle$ is the usual inner product.

A. Settling the subspace-security constraint

In this work, we are focusing on the subspace-security class of the WOA framework. Following Kerckhoffs' principle, an adversary is deemed to have access to all public data of the watermarking scheme: transformation domain where the embedding takes place (*e.g.* DCT, DWT) and can therefore access the feature vectors of size N_v of the contents. For the definition of the secret key, we use the same formalism as in [21], [34]: a secret key K_s generates a set of N_h secret components, $\mathcal{C} = \{\mathbf{c}_i\}_{i\in N_h}$. A message $\mathbf{m} \in \{0,1\}^{N_c}$ is associated to one or more secret components, meaning that $N_h \geq 2^{N_c}$.

We denote:

- $\mathbf{X} = (\mathbf{x}_0 \dots \mathbf{x}_{N_o-1})$: the set of N_o host original signals of size N_v (extracted from N_o host contents),
- $\mathbf{M} = (\mathbf{m}_0 \dots \mathbf{m}_{N_o-1})$: the set of N_o binary messages of size N_c (in the WOA framework, we assume that the digits of the messages are chosen at random),

- $\mathbf{Y}_{K_s} = (\mathbf{y}_0 \dots \mathbf{y}_{N_o-1})$: the set of N_o signals marked with the secret key K_s. Each signal $\mathbf{y}_j$ hides the message $\mathbf{m}_j$ on $\mathbf{x}_j$ with $j \in [N_o]$,
- $\mathcal{S}$: the set of embedding and decoding keys.

To efficiently modify the embedded message, the first step for an adversary is to estimate the secret components of K_s. According to Kerckhoffs' principle, he has access to:

- $\mathrm{p}(\mathbf{X})$: the distribution of host feature vectors which can be explicitly known (analytic formula) or modeled (the adversary can generate his own signals with contents he may have collected elsewhere),
- $\mathrm{p}(\mathbf{Y}|K_i)$: the conditional distribution of marked signals given a key $K_i \in \mathcal{S}$, K_i being generated by the adversary,
- $\mathrm{p}\left(\mathbf{Y}_{K_s}\right)$: the distribution of signals marked by the distributor with the secret key K_s, the adversary has access to the marked contents without the knowledge of secret key K_s.

A watermarking scheme qualifies for subspace-security if the distribution of marked signals is the same for each possible key. More precisely, the pdf $\mathrm{p}(\mathbf{Y}|K_i)$ does not depend on the key K_i. This assumption is equivalent to state that $\mathbf{Y}$ and K_i are independent. Formally [21]:

$$\forall K_i \in \mathcal{S},\ \mathrm{p}\left(\mathbf{Y}|K_i\right) = \mathrm{p}\left(\mathbf{Y}_{K_s}\right). \tag{1}$$

In this case, an adversary cannot estimate K_s because he sees no differences by trying several secret keys in $\mathcal{S}$. One way for a distributor to achieve subspace-security in WOA framework is to ensure that the distribution of the watermarked signals is identical to the distribution of the original host signals.

Note that this condition implies that the watermarking scheme also belongs to the key-security class. It means that there exists a subset of keys $\mathcal{S}_c \subset \mathcal{S}$ where distributions of marked signals are the same. An adversary can estimate this subset of keys which contains the secret key K_s:

$$\forall K_i \in \mathcal{S}_c,\ \mathrm{p}\left(\mathbf{Y}|K_i\right) = \mathrm{p}\left(\mathbf{Y}_{K_s}\right). \tag{2}$$

Finally, a watermarking scheme belongs to the insecurity class if for each secret key, the distribution of marked contents is unique, formally:

$$\exists!\ K_i \in \mathcal{S},\ \mathrm{p}\left(\mathbf{Y}|K_i\right) = \mathrm{p}\left(\mathbf{Y}_{K_s}\right), \tag{3}$$

and:

$$\forall K_1 \neq K_2 \in \mathcal{S},\ \mathrm{p}\left(\mathbf{Y}|K_1\right) \neq \mathrm{p}\left(\mathbf{Y}|K_2\right). \tag{4}$$

We now review some previous relevant works on secure WOA watermarking.

1) χ^2 Watermarking: In [35], we have proposed an illustrative subspace-secure scheme called χ^2 Watermarking (χ^2W). This technique modifies the square Euclidean norm of a host signal to embed a message, and the secret components are then partitioned along the positive real axis $\mathbb{R}^+$ according to a secret key. In fact, if $\mathbf{x} \in \mathbb{R}^{N_v}$ with $\mathbf{x} \sim \mathcal{N}(0,1)$, $\|\mathbf{x}\|^2 \sim \chi^2(N_v)$ (χ^2 law of degrees N_v). The secret components are then set in a partition of $[0, +\infty[$. To embed a secret message $\mathbf{m}$ in a host vector $\mathbf{x}$, we randomly choose a square norm $\mathrm{N}(\mathbf{m})$ in the corresponding real interval and we compute:

$$\mathbf{y} = \sqrt{\frac{\mathrm{N}(\mathbf{m})}{\|\mathbf{x}\|^2}}\mathbf{x}. \tag{5}$$

Watermarking is then done by selecting a square norm in the secret component corresponding to the message. By keeping the distribution of the square norms after watermarking, this technique is subspace-secure. Fig. 2 shows the distribution of the norms of $2,000$ watermarked signals $\|\mathbf{y}_j\|^2$ with messages of $N_c = 2$ bits.

However, a severe drawback of χ^2W is its weak robustness against white Gaussian noise addition, contrary to other methods such as spread-spectrum modulations.

2) Spread-spectrum Watermarking: Spread-spectrum techniques are good candidates for robust watermarking, which can also be extended to secure versions. We consider a message $\mathbf{m} \in \{0,1\}^{N_c}$ to be embedded into a host Gaussian distributed content $\mathbf{x}_j \in \mathbb{R}^{N_v}$. The message is coded using N_c carriers $\left\{\mathbf{u}_i \in \mathbb{R}^{N_v}\right\}_{i \in N_c}$ which are the secret components. These carriers are generated with a pseudo random number generator (PRNG) initialized with a seed $K_s \in \mathbb{N}$. They come as zero-mean Gaussian vectors obtained with the PRNG and are further orthogonalized to provide a basis of the private subspace (Sec. II-A4 provides more details about the generation of the carriers). In the WOA framework, security attacks are connected with the estimation of the carriers $\mathbf{u}_i$. It is not necessary to go back to the PRNG key K_s to perform a security attack. For the rest of this article, we make no difference between K_s and the carriers $\mathbf{u}_i$.

For each host signal $\mathbf{x}_j$, the embedding function e_{K_s} (depending on the secret key K_s) creates the marked signal $\mathbf{y}_j$ (in order to hide the message $\mathbf{m}_j$) following:

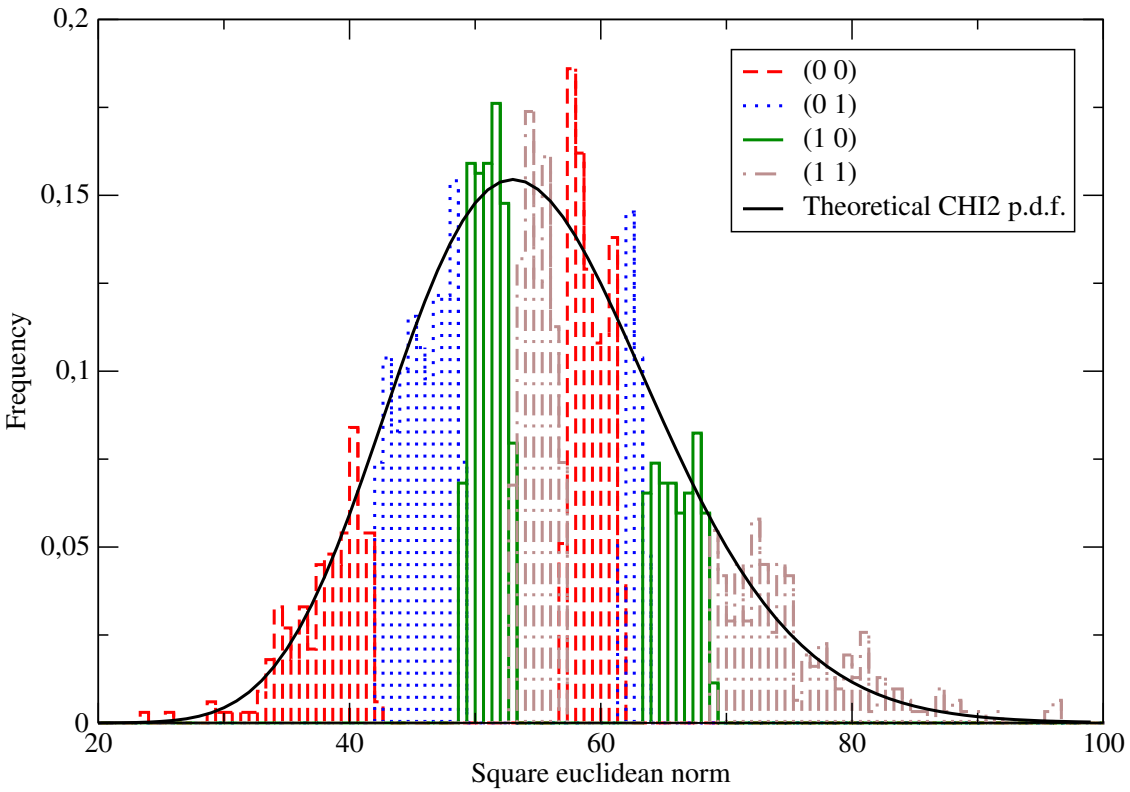


Figure 2. Distribution of square Euclidean norms of watermarked signals. We keep the $\chi^2(N_v)$ distribution of host norms. Parameters: $N_o = 2,000,\ N_c = 2,\ N_v = 55$.

$$\begin{aligned} \forall j \in [N_o],\ \mathbf{y}_j &= \mathrm{e}_{K_s}(\mathbf{x}_j, \mathbf{m}_j) \\ &= \mathbf{x}_j + \mathbf{w}_j \\ &= \mathbf{x}_j + \sum_{i=0}^{N_c-1} \mathrm{s}(\mathbf{m}_j(i), \mathbf{x}_j)\mathbf{u}_i, \end{aligned} \tag{6}$$

where $\mathbf{w}_j$ is the watermark signal and $\mathrm{s}:\ \{0,1\} \times \mathbb{R}^{N_v} \to \mathbb{R}$ is a modulation. Decoding is performed using correlations $\tilde{\mathbf{y}}_j \in \mathbb{R}^{N_c}$:

$$\forall i \in [N_c],\ \tilde{\mathbf{y}}_j(i) = \frac{1}{N_v}\langle \mathbf{y}_j | \mathbf{u}_i \rangle. \tag{7}$$

We obtain:

$$\hat{\mathbf{m}}_j(i) = \begin{cases} 0 & \text{if } \tilde{\mathbf{y}}_j(i) > 0, \\ 1 & \text{if } \tilde{\mathbf{y}}_j(i) \leq 0, \end{cases} \tag{8}$$

where $\hat{\mathbf{m}}_j(i)$ is the i-th decoded bit.

The classical Spread-Spectrum modulation (SS) is defined by:

$$\mathrm{s}_{SS}(\mathbf{m}_j(i), \mathbf{x}_j) = \alpha(-1)^{\mathbf{m}_j(i)}, \tag{9}$$

where α is a parameter driving the strength of the embedding, *i.e.* the distortion.

This modulation is known to belong to the insecure watermarking class [21], as secret carriers can be precisely estimated using first Principal Component Analysis (PCA) to estimate the secret subspace then Independent Component Analysis (ICA) [25], [36] to estimate the secret components (see II-A3). Subspace-secure spread-spectrum watermarking has been proposed in [28] with the Natural Watermarking (NW), this modulation enables to keep the same Gaussian distribution of correlations between host signals and secret carriers.

The s_{NW} modulation is given by:

$$\mathrm{s}_{NW}(\mathbf{m}_j(i), \mathbf{x}_j) = \Big((-1)^{\mathbf{m}_j(i)}\mathrm{sign}(\tilde{\mathbf{x}}_j(i)) - 1\Big)\,\tilde{\mathbf{x}}_j(i). \tag{10}$$

This technique uses symmetries of host correlations between the N_c secret carriers to embed the N_c bits with classical spread-spectrum decoding rules by keeping the relation $|\tilde{\mathbf{y}}_j(i)| = |\tilde{\mathbf{x}}_j(i)|$.

3) Practical security analysis: Considering that an adversary has access to N_o spread-spectrum watermarked contents $\mathbf{Y}_{K_s}$, his goal is to obtain as much information as possible about the secret components $\mathbf{U} = (\mathbf{u}_0 \dots \mathbf{u}_{N_c-1})$. Principal Component Analysis allows for an adversary to estimate the N_c-dimensional private subspace spanned by the secret components if the embedding alters the covariance matrix of the contents. This technique aims at finding the optimal linear transformation corresponding to the subspace yielding the largest variance. For classical SS modulation (Eq. (9)), message embedding increases the variance of the signal in the direction of the carriers. Note that PCA deals with subspace-security: contrary to classical SS technique, this technique cannot be correctly applied to contents watermarked with theoretical NW modulation. In Sec. III-D, we analyse the security of NW modulation used in a practical implementation (for still images) using PCA.

To measure the accuracy of the estimation of the private subspace, we use the chordal distance [14], [37], which provides a distance between two subspaces. If $\hat{\mathbf{U}} = (\hat{\mathbf{u}}_0 \dots \hat{\mathbf{u}}_{N_c-1})$ denotes the estimated carriers, the chordal distance between $\mathbf{U}$ and $\hat{\mathbf{U}}$ is defined by:

$$\mathrm{d}_c = \frac{1}{\sqrt{N_c}} \left(\sum_{i=0}^{N_c-1} \sin^2(\theta_i) \right)^{1/2}, \tag{11}$$

where the $\theta_0 \dots \theta_{N_c-1}$ denote the principal angles between $\mathbf{U}$ and $\hat{\mathbf{U}}$. The chordal distance d_c equals 0 when the subspaces spanned by the both matrices are the same and 1 when the matrices are orthogonal.

4) On the generation of the secret carriers: While there is no explicit agreement in the watermarking community on how to generate secret carriers, we describe here how we precisely generate ours:

- The Mersenne Twister MT19937 [38] PRNG is seeded with the secret key (this PRNG has no known bias, generates high-quality output for simulations and accepts seeds as long as 768 bits). We obtain uniformly distributed vectors;
- Going from a uniform to a Gaussian distribution is done with the Ziggurat method [39]. This method is more reliable and quicker than the usual Box-Muller transform [40];
- In order to remove inter-symbol interference, we explicitly orthogonalize the vectors with the help of the Gram-Schmidt procedure (we have verified that this procedure has a very limited impact, if any, on the gaussianity of the vectors). Further, we also explicitly remove the mean and normalize the output to obtain $\mathcal{N}(0,1)$-distributed vectors. These vectors are now the secret carriers we use.

In itself, the MT19937 is not a cryptographically secure PRNG. It is vulnerable to a Berlekamp-Massey attack [41], [42]. However, performing such an attack in the context of digital watermarking has not been done before and would prove extremely difficult because:

- An adversary would need to recover the exact, binary output of the MT19937 to perform the attack;
- He could only estimate the normalized, zero-mean output of the Ziggurat algorithm. To the best of our knowledge, the Ziggurat algorithm is not invertible and even the quality of the adversary's estimation may be insufficient in practice (when such an estimation is possible, *i.e.* when the embedding technique at hand is insecure). Also, the adversary would need to guess the original mean and variance of the vectors after applying the Ziggurat algorithm, all of which is lost information after the final step.

However, in case such an attack would prove feasible in practice, which we doubt, it would be sufficient to change for a cryptographically secure PRNG based, *e.g.*, on a standard cryptosystem. Thus, carriers estimation is the only attack one should consider in the context of digital watermarking security.

B. Optimization of imperceptibility

For the two subspace-secure techniques χ^2W and NW we have presented, in order to keep the distribution of the correlations after watermarking, it is not possible to change the distortion after embedding. Security of watermarking schemes in WOA framework can be completely defined by the distribution of signals after embedding (distribution of correlations for spread-spectrum techniques, distribution of Euclidean norms for χ^2W), and this property therefore constraints the distortion.

In this Section, watermarking in WOA is presented from a new perspective: we first generate a marked distribution in the subspace spanned by the secret key (a bijection between the security level one wants to achieve and the distribution). Next, we create marked signals by retro-projecting into the N_v-dimensional space.

From the distortion point of view, the subspace-secure techniques presented above are not optimal. In the WOA framework, we want to watermark N_o host contents with random messages. For each host content (a point in the host distribution in the private subspace), we associate a marked content in the targeted distribution. The question now becomes: how to associate each point of a host distribution to each point of a marked distribution in a decoding region by minimizing the distances between these points (the distance being proportional to the distortion induced by the watermark)?

1) The Hungarian optimization algorithm: In [35], we have proposed a way to match a host distribution in the private subspace with a marked distribution, while minimizing the global distance using the Hungarian algorithm [43]. This method is discrete and it is based on a precomputed matching on N_m host correlations ($N_m \geq N_o$) vectors $\tilde{\mathcal{X}}$ and N_m marked correlations vectors $\tilde{\mathcal{Y}}$. The Hungarian algorithm is able to find the bijection H which minimizes the Euclidean distance (weight) on average between $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$. To watermark a signal $\mathbf{x}$, we find the nearest neighbor of $\tilde{\mathbf{x}}$ in $\tilde{\mathcal{X}}$ and we obtain $\tilde{\mathbf{y}}$ based on the optimal matching given by the Hungarian algorithm. However, this method has three disadvantages. First, the complexity is $O({N_m}^3)$ and can be a real computational problem when N_m and N_o are important. Secondly, the performances (gain of distortion) decrease when the dimension of the private subspace (the number of secret components for SS techniques) increases, this is due to the difficulty to find Nearest Neighbors in high dimensional spaces. Finally, this method is bounded by the number of observations we use to compute the matching and it becomes largely sub-optimal for points lying beyond the tail of the empirical distribution (see Fig. 5). Fig. 3 shows an illustration of two subsets $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$ ($N_m = 12$), as well as the optimal matching H found by the Hungarian algorithm (weights are not depicted).

To avoid these problems, we present in the next subsection a method based on the transportation

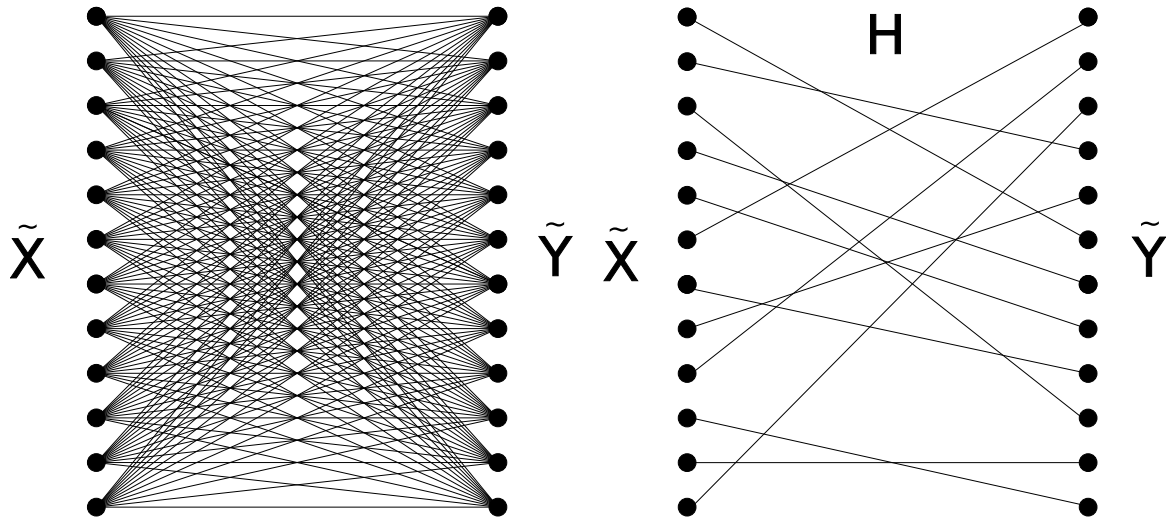


Figure 3. Example of triplet $(\tilde{\mathcal{X}},\ \tilde{\mathcal{Y}},\ \mathrm{H})$ with $N_m = 12$. H is the bijection obtained by the Hungarian method which minimizes the sum of weights of edges with complexity $\mathcal{O}({N_m}^3)$.

theory which minimizes the global distortion and its application on subspace-secure spread-spectrum watermarking.

2) Transportation theory: This theory has been first considered by mathematicians Monge [44] and Kantorovitch [45], [46]. This domain consists in finding a bijection T^* (a transport map) in order to move an initial set of points $\tilde{\mathcal{X}} \subset \mathbb{R}^N$ with distribution μ to a final set $\tilde{\mathcal{Y}} \subset \mathbb{R}^N$ with distribution ν given a cost function $\mathrm{c}:\ \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}} \to \mathbb{R}^+$. T^* is an optimal transport map if it minimizes the total cost as illustrated in Fig. 4. Formally, the problem is to find:

$$\mathrm{T}^* = \arg\min_{\mathrm{T}} \left\{ \int_{\mathbb{R}^N} \mathrm{c}(\tilde{\mathbf{x}}, \mathrm{T}(\tilde{\mathbf{x}}))\mu(\tilde{\mathbf{x}})\mathrm{d}\tilde{\mathbf{x}} \,\middle|\, \nu = \mu \circ \mathrm{T}^{-1} \right\}. \tag{12}$$

In [30], [47], if h, defined by $\mathrm{h}(\tilde{\mathbf{x}} - \tilde{\mathbf{y}}) = \mathrm{c}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$, is a convex function, a one-dimensional solution ($N = 1$) is given by:

$$\mathrm{T}^* = \mathrm{P}_\nu^{-1} \circ \mathrm{P}_\mu. \tag{13}$$

For $N \geq 1$, we can apply the Knott-Smith criterion [48]: if $\mathrm{c}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = \|\tilde{\mathbf{x}} - \tilde{\mathbf{y}}\|^2$, the sufficient conditions implying that T^* is a minimizer of Eq. (12) are:

i) $\mathrm{T}^*(\tilde{\mathcal{X}}) \sim \nu$ (image of $\tilde{\mathcal{X}}$ by T^* is distributed according to ν),

ii) the Jacobian matrix $\mathbf{J}_{\mathrm{T}^*}$ of T^* is symmetric and positive semi-definite.

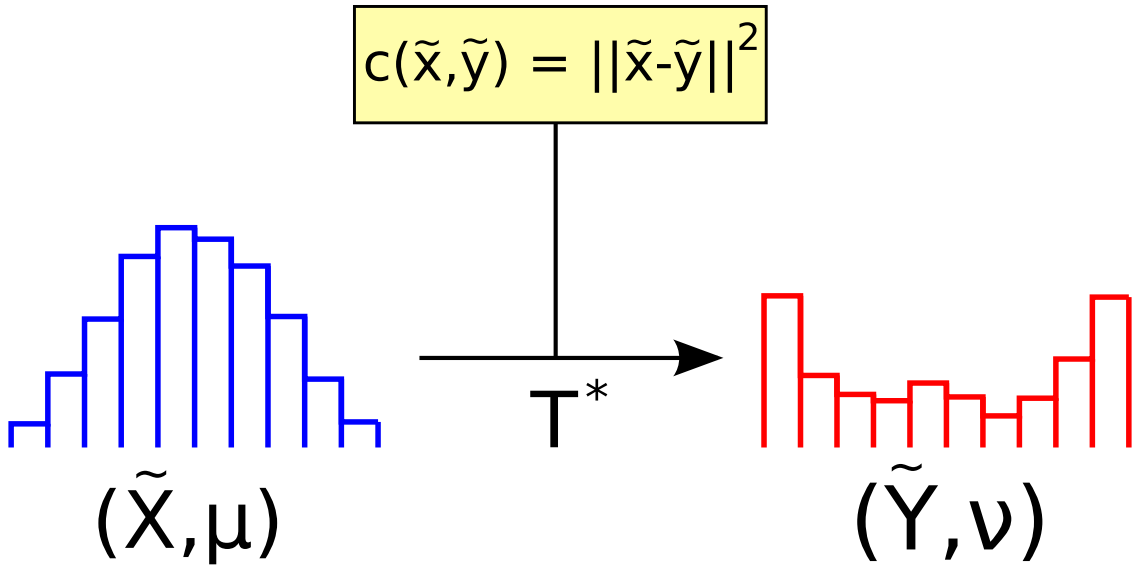


Figure 4. Illustration of the transportation problem: we look for an optimal transport map $\mathrm{T}^* : \tilde{\mathcal{X}} \to \tilde{\mathcal{Y}}$ with $\tilde{\mathcal{X}} \sim \mu$ and $\tilde{\mathcal{Y}} \sim \nu$ which minimizes the cost $\mathrm{c} : \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}} \to \mathbb{R}^+$.

One way to achieve subspace-security in spread-spectrum watermarking is to keep the distribution of the correlations of the contents over the N_c secret carriers. One can see the links between transportation theory and secure watermarking. Given a host distribution, it is possible to generate a marked distribution while minimizing the distortion cost. In our problem, we consider the minimization of the quadratic distance, in order to maximize the PSNR for instance, hence we set $\mathrm{c}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = \|\tilde{\mathbf{x}} - \tilde{\mathbf{y}}\|^2$.

3) Optimal embedding: We denote with $\tilde{\mathcal{X}} = \{\tilde{\mathbf{x}}_j\}_{j \in N_o} \subset \mathbb{R}^{N_c}$ the set of host correlations and $\tilde{\mathcal{Y}} = \{\tilde{\mathbf{y}}_j\}_{j \in N_o} \subset \mathbb{R}^{N_c}$ is the set of watermarked correlations. Further, we assume that host contents $\mathbf{x}$ are Gaussian distributed. This assumption is plausible due to the central limit theorem and to the fact that we work on host components projected on random vectors. We have therefore:

$$\forall i \in [N_c],\ \tilde{\mathbf{x}}(i) \sim \mathcal{N}\left(0, \sigma_{\mathbf{x}}^2 / N_v\right) = \mu, \tag{14}$$

with:

$$\mathrm{P}_\mu(t) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{t\sqrt{N_v}}{\sigma_{\mathbf{x}}\sqrt{2}}\right)\right). \tag{15}$$

Since spread-spectrum correlations are symmetric in the private subspace, we first consider the embedding of a N_c-bit constant message for each host content, for example $\mathbf{m} = \mathbf{0} = (0 \ldots 0)$, whose decoding

region is represented by the region $\mathbb{R}^{+^{N_c}}$ in the private subspace. To guarantee subspace-security, we need:

$$\forall i \in [N_c],\ \tilde{\mathbf{y}}(i) \sim \mathcal{N}^{+}(0, \sigma_{\mathbf{x}}^2/N_v) = \nu, \tag{16}$$

where $\mathcal{N}^{+}$ denotes a Gaussian distribution truncated in $\mathbb{R}^{+}$.

If $\delta \sim \mathcal{N}\left(0, \sigma_{\mathbf{x}}^2/N_v\right)$, $\nu \sim \mathcal{N}^{+}(0, \sigma_{\mathbf{x}}^2/N_v)$ is given by its cdf:

$$\mathrm{P}_{\nu}(t) = \begin{cases} 0 & \text{if } t < 0, \\ 2\mathrm{P}_{\delta}(t) - 1 & \text{if } t \geq 0. \end{cases} \tag{17}$$

The quantile function of P_{ν} is given by:

$$\mathrm{P}_{\nu}^{-1}(t) = \mathrm{P}_{\delta}^{-1}(t/2 + 1/2) = \frac{\sigma_{\mathbf{x}}\sqrt{2}}{\sqrt{N_v}}\mathrm{erf}^{-1}(t). \tag{18}$$

The strategy we adopt here is the following: we apply an optimal transport map for each dimension of the N_c-dimensional subspace using Eq. (13), Eq. (15) and Eq. (18). We consequently have the transport map $\mathrm{T}_{\mathbf{0}} : \mathbb{R}^{N_c} \to \mathbb{R}^{N_c}$:

$$\tilde{\mathbf{y}} = \mathrm{T}_{\mathbf{0}}\left(\tilde{\mathbf{x}}\right) = \begin{pmatrix} \mathrm{P}_{\nu}^{-1} \circ \mathrm{P}_{\mu}(\tilde{\mathbf{x}}(0)) \\ \vdots \\ \mathrm{P}_{\nu}^{-1} \circ \mathrm{P}_{\mu}(\tilde{\mathbf{x}}(N_c - 1)) \end{pmatrix}. \tag{19}$$

Proposition 1. $\mathrm{T}_{\mathbf{0}}$ *is an optimal transport map.*

Proof:

This proposition is intuitively straightforward due to the separability of the problem along the N_c orthogonal secret components. However, in order to prove it rigorously, we have to use the Knott-Smith criterion (Sec. II-B2) to prove that $\mathrm{T}_{\mathbf{0}}$ is an optimal transport map. Assumption i) is verified because $\tilde{\mathbf{y}}$ is a Gaussian vector (its components are Gaussian i.i.d.). In order to verify the assumption ii), we introduce the Jacobian matrix of $\mathrm{T}_{\mathbf{0}}$:

$$\mathbf{J}_{\mathrm{T_o}}$$

$$= \begin{pmatrix} \frac{\partial(\mathrm{P}_\nu^{-1}\circ\mathrm{P}_\mu(\tilde{\mathbf{x}}(0)))}{\partial\tilde{\mathbf{x}}(0)} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{\partial(\mathrm{P}_\nu^{-1}\circ\mathrm{P}_\mu(\tilde{\mathbf{x}}(N_c-1)))}{\partial\tilde{\mathbf{x}}(N_c-1)} \end{pmatrix}. \tag{20}$$

We now prove that this matrix is positive semi-definite. By construction of $\mathrm{T}_{\mathbf{0}}$, $\mathbf{J}_{\mathrm{T_o}}$ is symmetric. Then $\mathbf{J}_{\mathrm{T_o}}$ is positive semi-definite iff eigenvalues of $\mathbf{J}_{\mathrm{T_o}}$ are positives or null. We now have to prove that $\mathrm{P}_\nu^{-1}\circ\mathrm{P}_\mu(t) \geq 0$. We have:

$$\left(\mathrm{P}_\nu^{-1}\circ\mathrm{P}_\mu\right)'(t) = \left(\mathrm{P}_\nu^{-1}\right)'(\mathrm{P}_\mu(t))\,\mathrm{f}_\mu(t). \tag{21}$$

f_μ is positive (pdf) and P_ν^{-1} is a non-decreasing function, so $\left(\mathrm{P}_\nu^{-1}\right)'(t) \geq 0$. Finally $\left(\mathrm{P}_\nu^{-1}\circ\mathrm{P}_\mu\right)'(t) \geq 0$ and $\mathbf{J}_{\mathrm{T_o}}$ is positive semi-definite: $\mathrm{T}_{\mathbf{0}}$ is an optimal transport map.

■

Because of the symmetry property of decoding with correlations in spread-spectrum watermarking, in order to embed messages which differ from $\mathbf{0}$, central symmetries have to be performed on components of $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ before and after embedding to correctly hide the message $\mathbf{m}$. For any $\mathbf{m} \in \mathbb{F}_2^{N_c}$, the optimal transport map $\mathrm{T}_{\mathbf{m}}$ is given by:

$$\mathrm{T}_{\mathbf{m}}(\tilde{\mathbf{x}}) = \mathbf{R}^{-1}(\mathrm{T}_0(\mathbf{R}\tilde{\mathbf{x}})), \tag{22}$$

with:

$$\mathbf{R} \in \mathcal{M}_{N_c,N_c}(\mathbb{R}), \mathbf{R}(i,j) = \begin{cases} 0 \text{ if } i \neq j, \\ (-1)^{\mathbf{m}(i)} \text{ otherwise.} \end{cases} \tag{23}$$

This optimal transport map allows us to design a new NW modulation called Transportation Natural Watermarking (TNW) where the modulation is given by:

$$\mathrm{s}_{TNW}(\mathbf{m}(i),\mathbf{x}) = \mathrm{T}_{\mathbf{m}}(\tilde{\mathbf{x}})(i) - \tilde{\mathbf{x}}(i). \tag{24}$$

We can also compute the distortion gain offered by TNW over NW, *e.g.* the difference between the two Watermark to Content Ratios $WCR = 10\log_{10}(\sigma_{\mathbf{w}}^2/\sigma_{\mathbf{x}}^2)$. Without loss of generality, we compute $\mathbb{E}\left[WCR\right]$ for $\mathrm{s} = \mathrm{s}_{NW}$ and $\mathrm{s} = \mathrm{s}_{TNW}$ considering $\mathbf{m}(i) = 0$.

For $\mathrm{s} = \mathrm{s}_{NW}$:

$$\begin{aligned}
\mathbb{E}\left[\mathrm{s}(\mathbf{m}(i),\mathbf{x})^2\right] &= \mathbb{E}\left[\left(\left(\tilde{\mathbf{x}}(i)/|\tilde{\mathbf{x}}(i)|-1\right)\tilde{\mathbf{x}}(i)\right)^2\right] \\
&= 1/2\times\mathbb{E}\left[4\tilde{\mathbf{x}}(i)^2\right] \\
&= 2/{N_v}^2\mathbb{E}\left[\langle\mathbf{x}|\mathbf{u}_i\rangle^2\right] \\
&= 2/{N_v}^2\mathbb{E}\left[\left(\sum_{j=0}^{N_v-1}\mathbf{x}(j)\mathbf{u}_i(j)\right)^2\right] \\
&= 2/{N_v}^2\times\sum_{j=0}^{N_v-1}\mathbb{E}\left[\mathbf{x}(j)^2\right]\times\mathbb{E}\left[\mathbf{u}_i(j)^2\right] \\
&= 2\sigma_{\mathbf{x}}^2/N_v.
\end{aligned} \tag{25}$$

We obtain then:

$$\mathbb{E}\left[WCR_{NW}\right] = 10\log_{10}\left(\frac{\sigma_{\mathbf{w}}^2}{\sigma_{\mathbf{x}}^2}\right) = 10\log_{10}\frac{2N_c}{N_v}. \tag{26}$$

For $\mathrm{s} = \mathrm{s}_{TNW}$:

$$\mathbb{E}\left[\mathrm{s}(\mathbf{m}(i),\mathbf{x})^2\right] = \int_{-\infty}^{+\infty}\left(\mathrm{P}_{\nu}^{-1}\circ\mathrm{P}_{\mu}(t)-t\right)^2\mathrm{f}_{\mu}(\mathrm{t})\mathrm{d}t. \tag{27}$$

$A = \mathbb{E}\left[\mathrm{s}(\mathbf{m}(i),\mathbf{x})^2\right]$ has to be computed using numerical integration. We obtain:

$$\mathbb{E}\left[WCR_{TNW}\right] = 10\log_{10}\left(\frac{\sigma_{\mathbf{w}}^2}{\sigma_{\mathbf{x}}^2}\right) = 10\log_{10}\frac{A\times N_c}{\sigma_{\mathbf{x}}^2}. \tag{28}$$

The WCR gain using TNW modulation instead of NW modulation is given by:

$$\begin{aligned}
Gain_{[dB]} &= \mathbb{E}\left[WCR_{NW}\right] - \mathbb{E}\left[WCR_{TNW}\right] \\
&= 10\log_{10}\left(\frac{2\sigma_{\mathbf{x}}^2}{AN_v}\right).
\end{aligned} \tag{29}$$

After performing numerical integration, we notice that $\mathbb{E}\left[WCR_{TNW}\right]$ is constant w.r.t $\sigma_{\mathbf{x}}^2$ and that the gain is constant w.r.t N_c. This gain is approximately ≈ 3.77 dB with $N_v = 256$.

Fig. 5 shows the embedding functions e (Eq. (6)) for NW, TNW and Hungarian HNW [35] modulations functions of scalar host signal $x \sim \mathcal{N}(0,1)$, where only one bit is embedded. We can see that, unlike classical modulation, improved methods are not linear. Moreover, because HNW is based on nearest neighbors on precomputed Gaussian mapping ($N_m = 1,000$ in this example), this method deals with edge effects (the min and max values) when x is increasing.

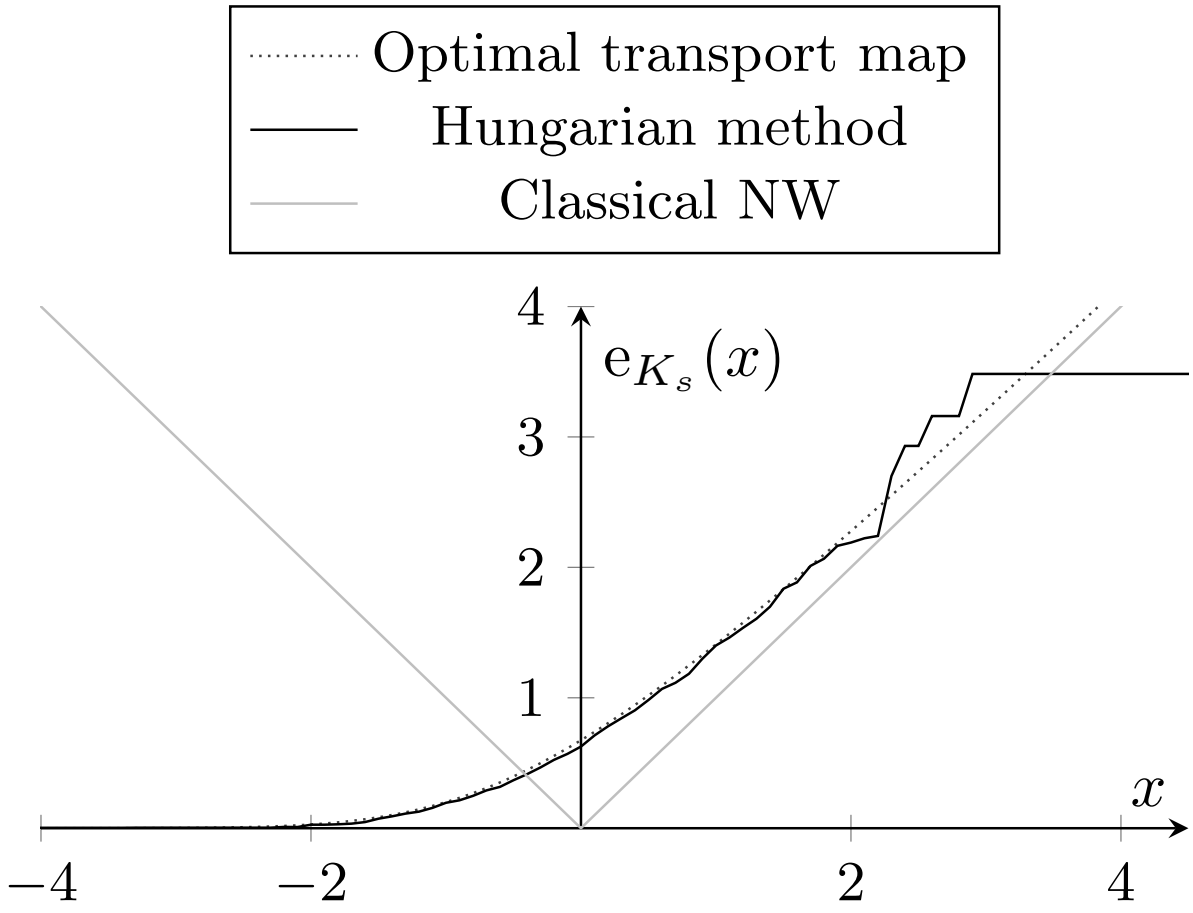


Figure 5. Embedding functions e with $N_v = 1$ and $\mathbf{m}(0) = 0$. The Hungarian algorithm is run with $N_m = 1,000$. We can see that the mapping provided by the Hungarian method is inaccurate for values above the maximum value of the empirical distribution.

Fig. 6 shows the distortion (measured by the WCR) caused by watermark embedding for NW, TNW and HNW modulations and its evolution w.r.t. the number of inserted bits N_c. As expected, the distortion gain between NW and TNW is constant w.r.t. the length of the embedded messages, these practical computations also match the theoretical derivations in (26) and (28) by verifying a constant gap. Additionally, we can notice that the distortion gap between NW and HNW decreases when the dimension of the private subspace (N_c) increases, this is due to the fact that the nearest neighbor search involves larger Euclidean distances and consequently incurs a loss of efficiency in high dimensional spaces.

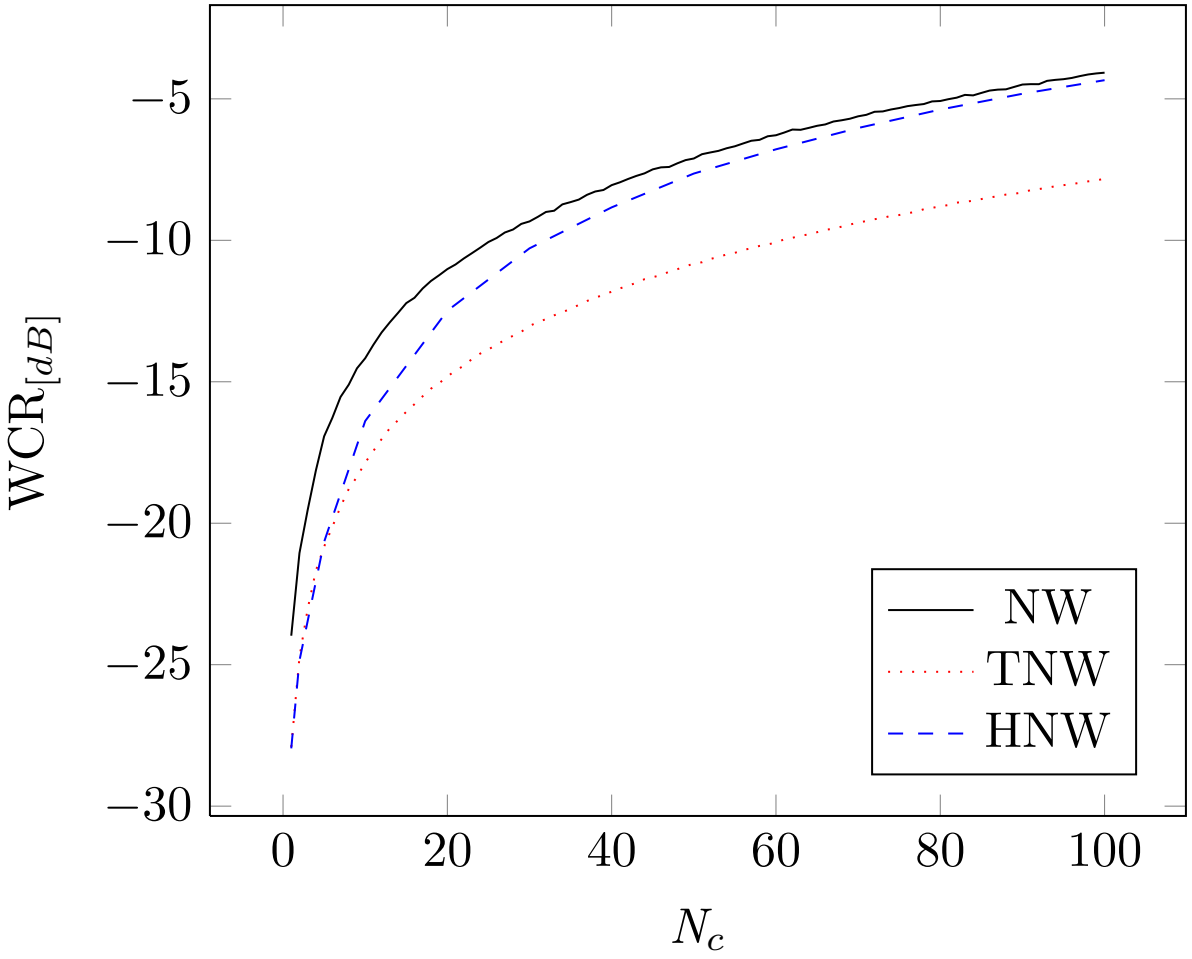


Figure 6. Distortion (WCR) w.r.t. the number of bits N_c for NW, HNW and TNW modulations, $N_v = 512$, $\sigma^2_{\mathbf{x}} = 1$. The Hungarian algorithm is computed with $N_m = 10,000$ Gaussian signals and the distortion is computed on average on $N_o = 2,000$ Gaussian signals for the three modulations. As mentioned in Sec. II-B1, the distortion difference between NW and HNW decreases when the dimension of the private subspace (N_c) increases because of larger Euclidean distances in large dimensions. On the contrary, TNW does not suffer from dimension issues and produces better results than HNW.

III. Application on still images

A. *From theoretical to practical secure watermarking*

Most watermarking schemes assume a Gaussian distribution of host signals. However, this model does not fit usual distributions of image components. For example, DCT components are often modeled by a Laplace distribution and DWT components by a generalized Gaussian distribution. In order to be close to the Gaussian assumption, we adopt the following strategy: we project the host feature vectors onto pseudo-random carriers generated with the same distribution. Following the Central Limit Theorem, the projected

signals (marginal distributions) are asymptotically Gaussian. By working on Gaussian distribution and applying NW, we can guarantee that we are subspace-secure in the projected subspace. If one is able to estimate the projected subspace, using denoising techniques for example [49], then the key-security property still holds because of the security of the Gaussian distribution created by the NW embedding.

B. Experimental proposed scheme

Fig. 7 presents our experimental watermarking scheme on grayscale images. After a 5-level 9/7 Daubechies DWT transform [50] on the host image, we select N_t components on 9 subbands in low/mean frequencies as presented in Fig. 8 to obtain a PSNR between original and marked images between 35 and 55 dB on average. Note that this DWT transform is not an orthogonal but a biorthogonal transform [51], however the 9/7 filter set deviates by only a few percent from orthogonal filter weighting [52]. The extracted feature vector with size N_t is denoted as $\mathbf{x_t}$. In order to respect a normal distribution on host contents to be able to apply TNW, we use a projection of the feature vector on N_v carriers $\{\mathbf{a}_i\}_{i\in N_v}$ with size N_t:

$$\forall i \in [N_v],\ \mathbf{x}(i) = \sum_{j=0}^{N_t-1} \mathbf{x_t}(j)\mathbf{a}_i(j). \tag{30}$$

Vectors $\mathbf{a}_i$ are pseudo-randomly generated according to an uniform distribution: $\forall i \in N_v,\ \forall j \in N_t,\ \mathbf{a}_i(j) \sim \mathcal{U}(-\sqrt{3/N_c}, \sqrt{3/N_c})$, quasi-orthogonal and $\mathbb{E}[||\mathbf{a}||^2] = 1$. For each $i \in N_v$, $\mathbf{x}(i)$ is asymptotically Gaussian distributed due to the CLT. Note that the independence between each component is not provided. However, in this work, we assume this condition, partially justified by the important length of the considered feature vectors.

The watermark signal $\mathbf{w}$ (Eq. (6)) is constructed by spread-spectrum TNW modulation for a message $\mathbf{m}$ and N_c secret carriers $\mathbf{u}_i$. To preserve the host distribution on the projected space and apply the TWN modulation (19), we first estimate values of $\sigma^2_{\mathbf{x}}$ for each selection of subbands on the whole BOWS2-IMAGES database [53] which contains $10,000$ images.

The watermark is then computed in the wavelet domain using the following inverse projection:

$$\forall i \in [N_t],\ \mathbf{w_t}(i) = \sum_{j=0}^{N_v-1} \mathbf{w}(j)\mathbf{a}_j(i). \tag{31}$$

In order to improve imperceptibility, we propose to add a psychovisual masking proposed by Piva [54] once the watermark is computed in the wavelet domain (Eq. (31)). Multiplicative embedding is consequently used to compute the final watermarked vector $\mathbf{y_t}$:

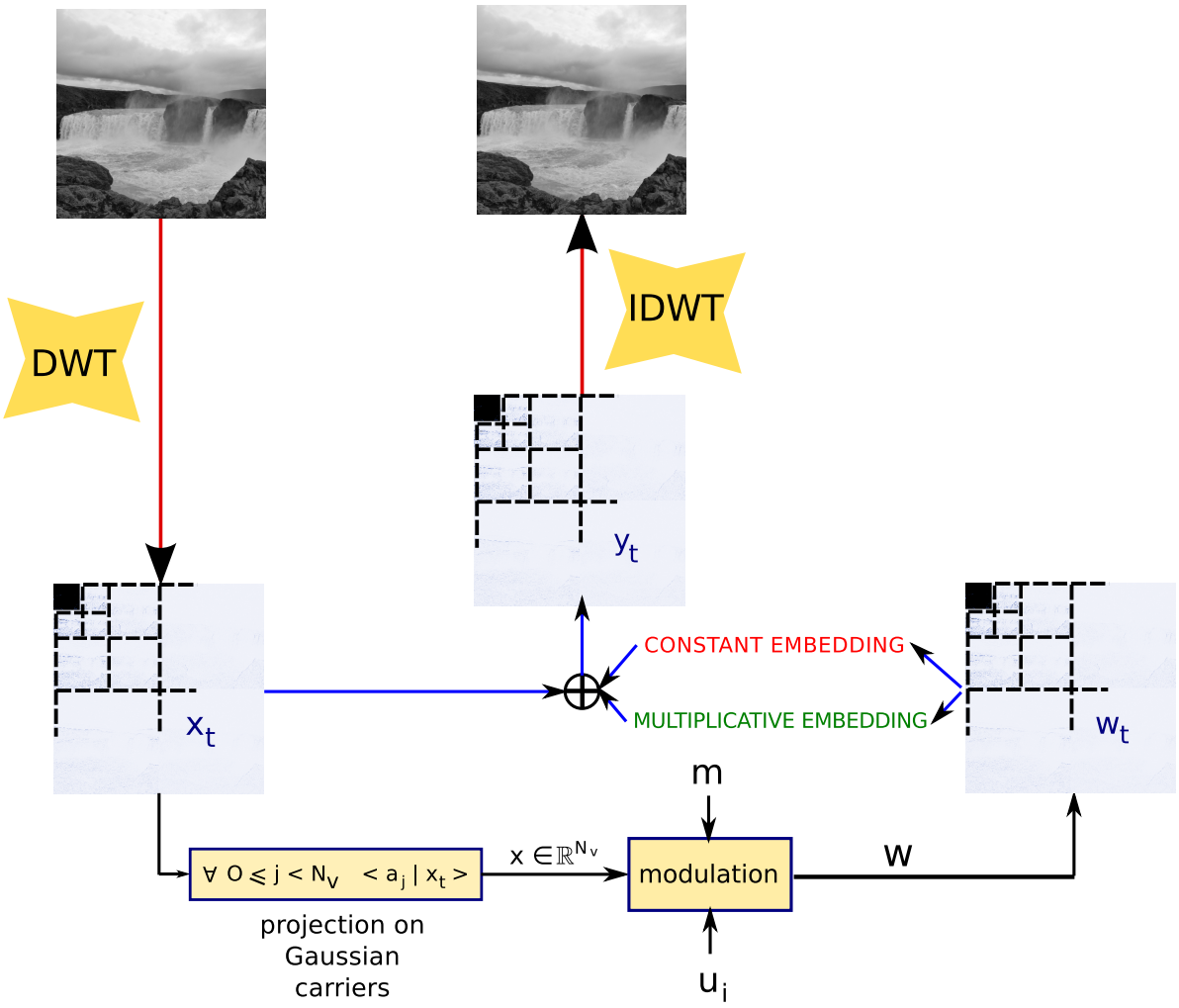


Figure 7. Experimental image watermarking scheme.

$$\mathbf{y_t} = \mathbf{x_t} + \mathbf{w'_t} \text{ with } \forall i \in [N_t],\ \mathbf{w'_t}(i) = \frac{1}{\mathbb{E}[|X_\mathbf{t}|]} |\mathbf{x_t}(i)| \mathbf{w_t}(i). \tag{32}$$

The normalization factor $(\mathbb{E}[|X_\mathbf{t}|])^{-1}$ guarantees that the distributions related to the projections $\langle \mathbf{w'_t} | \mathbf{a}_j \rangle$ and $\langle \mathbf{w_t} | \mathbf{a}_j \rangle$ remain identical after the psychovisual weighting ($\mathbb{E}[\langle W'_\mathbf{t} | A_j \rangle] = \mathbb{E}[\langle W_\mathbf{t} | A_j \rangle]$). Note that such a strategy does not change the optimality of the embedding w.r.t. the quadratic distance since in the secret subspace it is equivalent to the identity application.

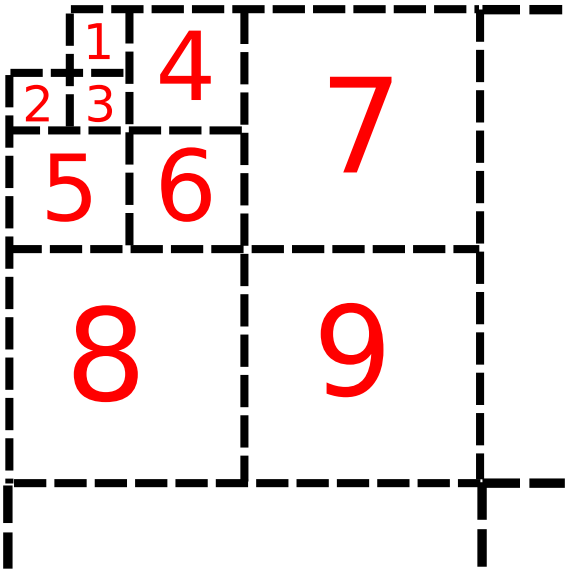


Figure 8. Component selection: the watermark is hidden on each possible selection using 9 subbands in low-frequencies components after a 9/7 Daubechies DWT with a 5-level decomposition in order to obtain a PSNR between 35 and 55 dB.

C. Subbands selection for distortion and decoding issues

Natural Watermarking (NW) and our improved Transportation Natural Watermarking (TNW) do not allow to set the embedding distortion in the N_v-dimensional subspace. However, the strength of the watermark can be tuned by selecting appropriate combinations of subbands in low-frequencies components. Fig. 8 depicts the 9 different components that can be selected to perform the embedding and we look for the selection which offers both the desired distortion and the best robustness. We experiment here our scheme on the $2^{9=}512$ possible combinations of subbands on $2,000$ 512×512 natural images from the BOWS2-IMAGES database [53]. We hide here $N_c = 16$ bits on each image and set the length of projected signals with $N_v = 256$. To test the robustness, we apply a JPEG compression with quality factor $Q_F = 30$ after watermarking and before decoding.

Fig. 9 shows the histogram of PSNR, on average, for the different combinations. The average PSNR is 44.05 dB with a deviation of 2.64 dB and the PSNRs range between 35 dB and 55 dB. Note that the relation between PSNR and WCR is given by [25]:

$$PSNR = 10 \log_{10} \left(\frac{255^2}{\sigma_{\mathbf{x}}^2} \times \frac{512^2}{N_v} \times \frac{\mathbb{E}[|X_{\mathbf{t}}|]^2}{\mathbb{E}[X_{\mathbf{t}}^2]} \right) - WCR. \tag{33}$$

Fig. 10 presents the relation between the PSNR computed over the original and watermarked images, and the Bit Error Rate after JPEG compression with $Q_F = 30$ for the 512 combinations. The couples

PSNR [dB]	BER	Combination
35.76	6.509375e-02	[1 2]
36.56	5.478125e-02	[1 2 3]
39.53	4.034375e-02	[1 2 3 4 5]
39.40	4.121875e-02	[1 2 3 4 5 6]
43.07	5.334375e-02	[4 5]
44.05	6.268750e-02	[4 5 6]
46.97	8.412500e-02	[4 5 6 7 8]
47.56	8.821875e-02	[4 5 6 7 8 9]
50.03	1.302812e-01	[7 8]
51.01	1.209375e-01	[7 8 9]

Table I

COUPLES (BER, PSNR) (IN CROSS PLOTS IN FIG. 10) SELECTED ACCORDING TO AN ITERATIVE PROCEDURE WHICH CONSISTS, IN A CONTINUOUS WAY, IN SELECTING HLi AND LHi SUBBANDS (AND HHi TO IMPROVE ROBUSTNESS) AND REMOVING LOW ONES UNTIL THE DESIRED DISTORTION IS ACHIEVED.

(BER, PSNR) which offer the lowest BER are depicted by square marks.

Several remarks can be outlined from these results:

1) The PSNR gain obtained using TNW instead of NW modulation is 3.6 dB on average with a standard deviation of 0.21 dB on the 512 combinations. This gain matches our theoretical computation using Eq. (29) and Eq. (33).
2) The lowest PSNRs correspond to low-frequencies subbands and the largest ones to mean-frequencies subbands.
3) The couples (BER, PSNR) with the lowest BERs correspond to combinations of contiguous or nearly contiguous sets of subbands. In order to increase the robustness for a given distortion budget, the best strategy is to pick the maximum number of low frequency subbands (in order to increase the spreading factor).
4) We can propose an iterative algorithm to select the combination which gives the lowest BER. It consists in selecting first subbands HLi and LHi corresponding to a distortion level, then adding the HHi subbands (components with a smaller variance) to improve robustness and, in a continuous way, adding mean-pass components and remove low ones to improve imperceptibility. Couples (BER, PSNR) for the proposed iterative algorithm are given in Tab. I and are depicted as crosses on Fig. 10.

Fig. 11 presents one of the $2,000$ images of 512×512 pixels watermarked with our experimental scheme

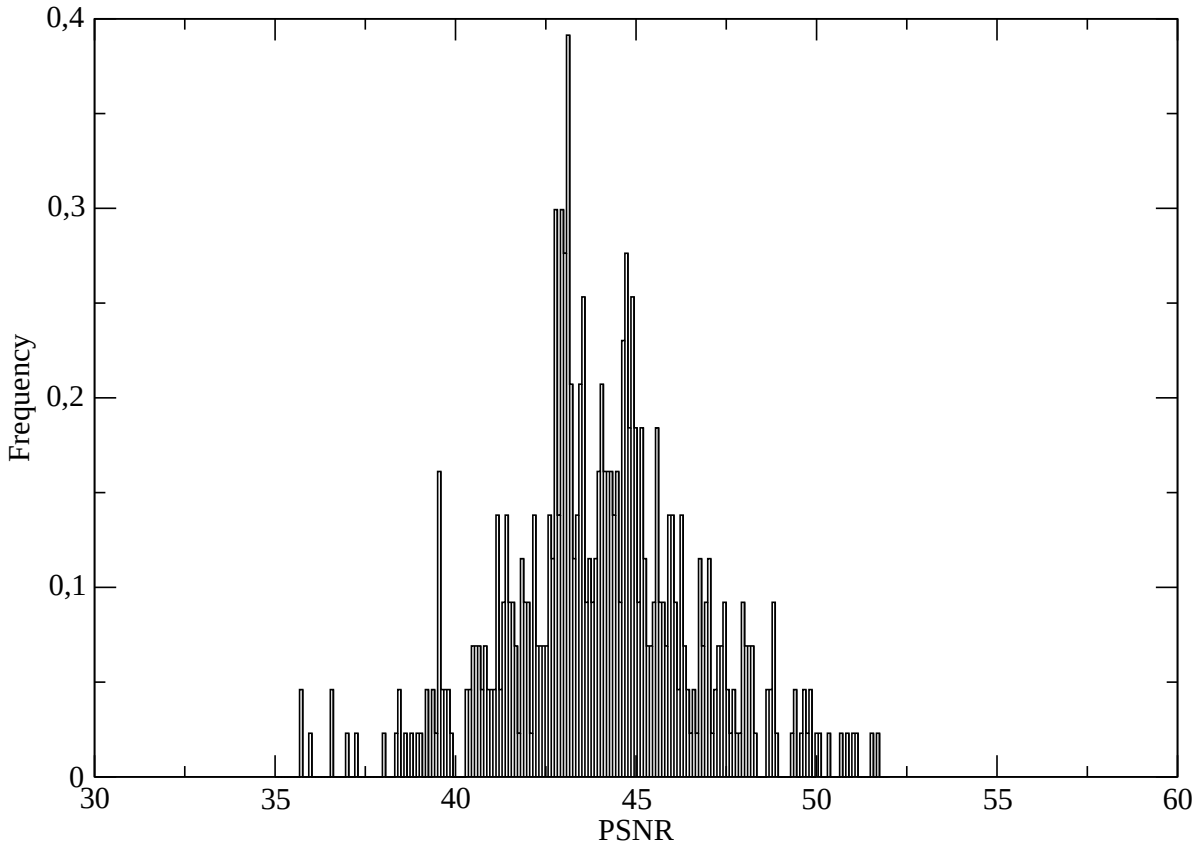


Figure 9. Histogram of PSNR for the 512 possible combinations of subbands. The mean is achieved by 44.05 dB and the standard deviation by 2.64 dB.

using the same [1 2] combination for NW and TNW modulations. Fig. 13 presents a zoom of upper areas of these images. We show here the consequences of using TNW instead of NW modulation and multiplicative embedding (Eq. (32)) instead of constant embedding (Eq. (31)) from the imperceptibility point of view.

Fig. 12 shows the robustness evaluation of our experimental watermarking scheme against JPEG compression: Bit Error Rate functions of JPEG Quality Factor (Q_F). We present here the four combinations which give the smallest BER after JPEG compression with $Q_F = 30$.

D. Security assessments

We assess the security analysis of the proposed scheme using two means:

1) Distributions: we perform visual inspection in the subspace spanned by the secret components in order to verify that the distribution after embedding is not modified. Fig. (14) shows the distribution of

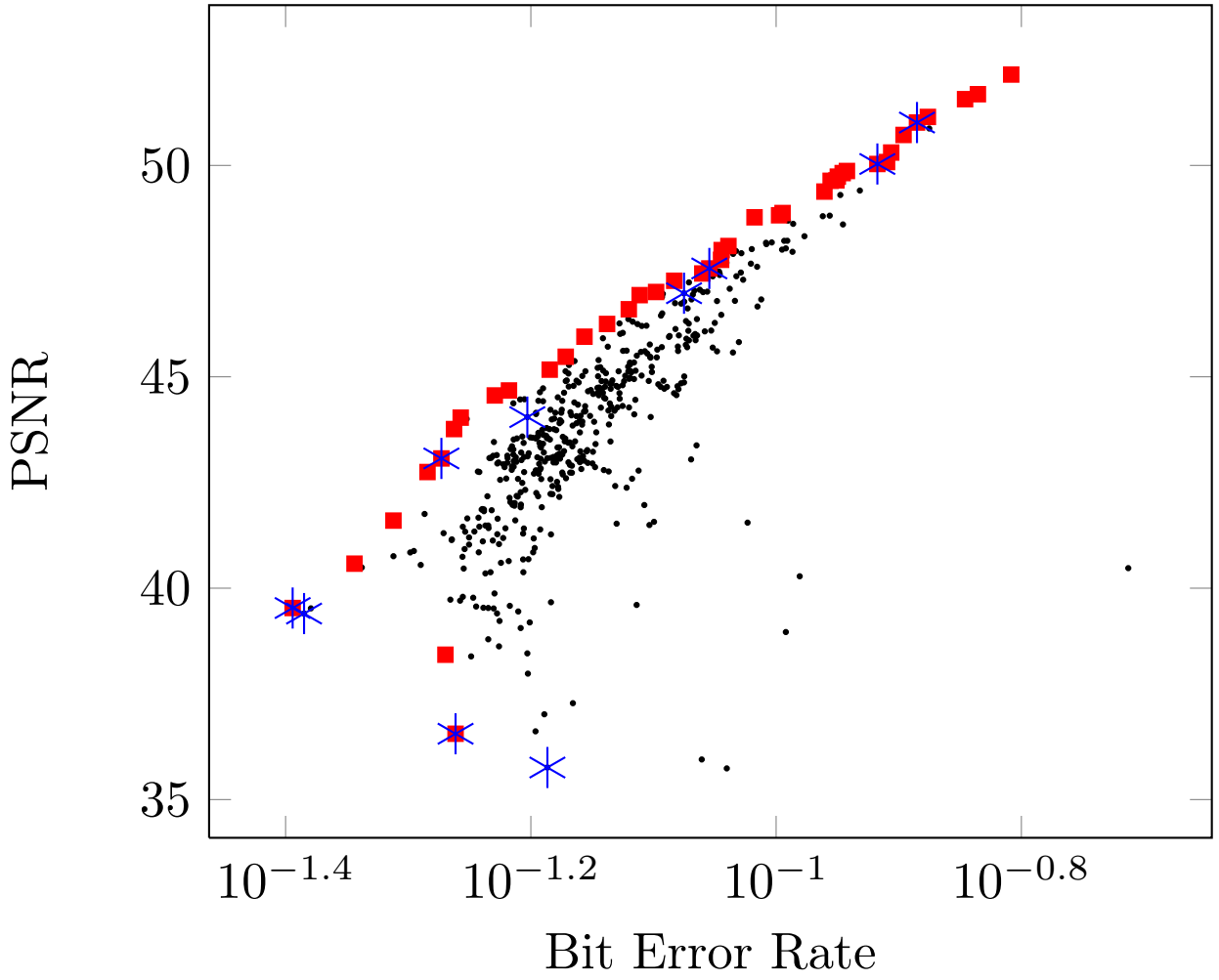


Figure 10. Average embedding PSNR w.r.t. the average BER after JPEG compression ($Q_F = 30$) on $2,000$ still images for $2^9 = 512$ possible combinations. Square plots show the best combinations (low BER) for PSNR between 35 and 55 dB and cross plots are selected according an iterative procedure on subbands as presented in Tab. I.

correlations between N_v-dimensional hosts and marked TNW signals using two secret carriers for the selection of subbands [3 4 5 6 7 8]. Distributions keep Gaussian after projection on quasi-orthogonal uniform carriers (Eq. (30)) and multiplicative embedding (Eq. (32)).

2) PCA: we try to estimate the secret subspace using PCA and measuring the distance between the estimated space and the true one using the chordal distance, then we compare our results with traditional, insecure SS. We apply here the method introduced in Sec. II-A3 It consists in estimating, using Principal Component Analysis (PCA) on the watermarked contents in $\mathbb{R}^{N_v}$, the subspace spanned by the secret components. Due to the embedding process, the variance of this subspace increases for insecure

(a) HOST

(b) NW (CE)

PSNR = 33.12 dB

(c) TNW (CE)

PSNR = 39.37 dB

(d) TNW (ME)

PSNR = 40.1 dB

Figure 11. Host image (a), Watermarked image using the NW modulation with constant embedding (CE) (b), the TNW modulation with CE (c) and the TNW modulation with multiplicative embedding (ME) (d). Parameters : combination = [1 2], $N_t = 512,\ N_v = 256,\ N_c = 16$.

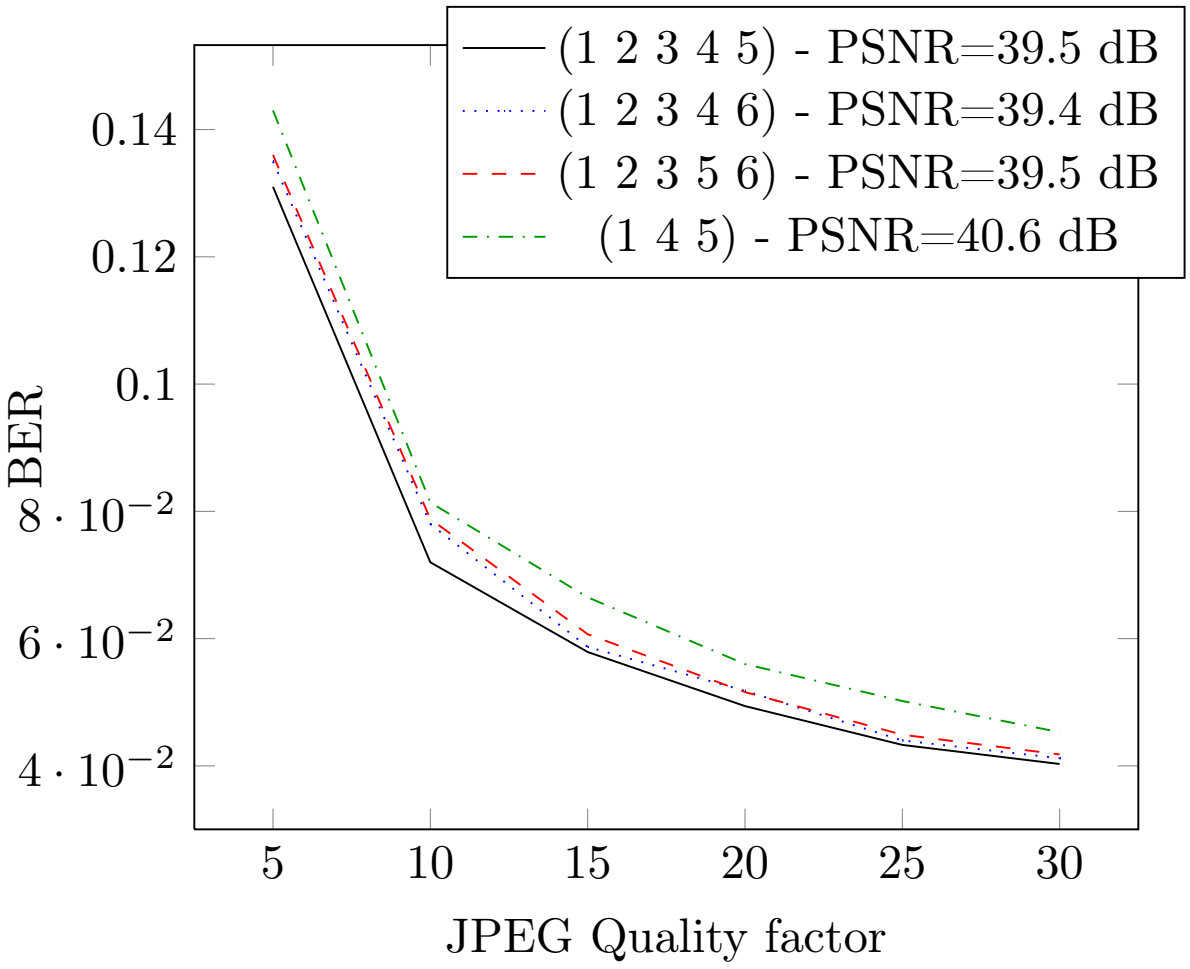


Figure 12. BER w.r.t. JPEG quality factor: the 4 combinations which give the smallest BER with a quality factor of 30.

embedding and consequently can be estimated by extracting the principal components. We consequently perform PCA on N_o watermarked N_v-dimensional signals from the $2,000$ images. Next we compute the chordal distance d_c (Eq. (11)) between the secret carriers $\mathbf{U}$ and N_c principal directions obtained after PCA $\hat{\mathbf{U}}$ of the N_c first eigenvectors (which corresponds to the N_c components with the largest variance). Signals are watermarked using multiplicative embedding with subbands selection [4 5 6 7 8 9] ($N_t = 15,360$, PSNR = 47.5 dB) and with TNW or classical spread-spectrum modulation (SS). For TNW, we obtain a chordal distance of 0.98 using the N_c first eigenvectors, which means that PCA is not able to correctly estimate the secret subspace. For SS, the chordal distance equals 0.08 when we focus on the N_c first eigenvectors which means that in this case the estimated and secret subspaces are very close. We can then conclude this analysis by stating that the proposed implementation is immune to

(a) HOST

(b) NW (CE) - PSNR = 33.12 dB

(c) TNW (CE) - PSNR = 39.37 dB

(d) TNW (ME) - PSNR = 40.1 dB

Figure 13. Zoom of upper areas of images presented in Fig. 11. Host image (a), Watermarked image using NW with constant embedding (CE) (b), TNW with CE (c) and TNW with multiplicative embedding (ME) (d). Parameters : combination = [1 2], $N_t = 512$, $N_v = 256$, $N_c = 16$.

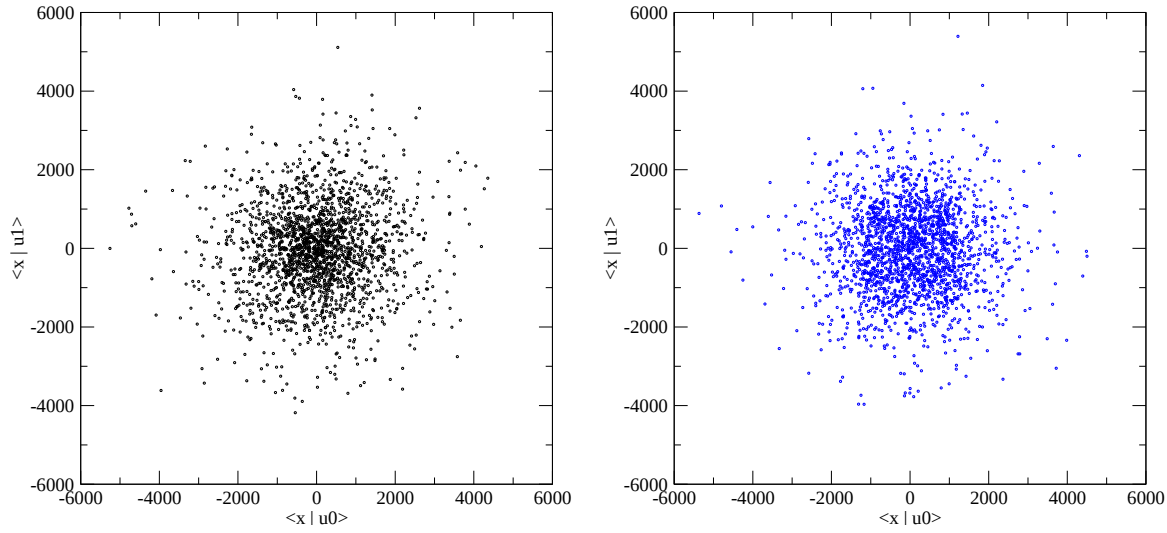


Figure 14. Distribution of correlations between two secret carriers and N_v-dimensional host (a) and watermarked (b) signals. Selected subbands: [3 4 5 6 7 8]. Parameters: $N_o = 2,000$, $N_t = 11,520$, $N_v = 256, N_c = 16$, PSNR $= 46.6$ dB. Distributions keep Gaussian after projection on quasi-orthogonal uniform carriers (Eq. (30)) and multiplicative embedding (Eq. (32)).

security attacks based on subspace estimation: this is due to the subspace-security property of the TNW embedding.

IV. Conclusion

This paper shows how to consider image watermarking by taking into account the three fundamental constraints: the security constraint is set to a given class (key or subspace-security), the distortion is minimized, and the robustness is maximized.

We have proposed to minimize the embedding distortion using both optimal transportation theory to guarantee the security class, and multiplicative embedding to minimize the visual impact. In the context of secure embedding, the robustness solely relies on the variance of the components of a given image and it is maximized by selecting the appropriate subbands. These two optimizations enable on one hand to obtain a PSNR gain of 3.6 dB on average, and on the other hand to provide the sets of configurations which maximize the robustness for a given distortion. Such a methodology can be also applied on other classes of secure watermarking schemes, by using optimal transport whenever it is possible [32], [33] or finding approximations of the optimal mappings for more complex distributions such as circular watermarking [29].

REFERENCES

[1] I. J. Cox, M. L. Miller, and A. L. McKellips, "Watermarking as communications with side information," *Proc. IEEE*, vol. 87, no. 7, pp. 1127–1141, Jul 1999.

[2] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proc. IEEE*, vol. 87, no. 7, pp. 1079–1107, Jul 1999.

[3] M. Barni and F. Bartolini, *Watermarking Systems Engineering: Enabling Digital Assets Security and Other Applications*. Marcel Dekker, 2004.

[4] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*, 2nd ed., Morgan Kaufmann Publishers In, Ed. The Morgan Kaufmann Series in Multimedia Information and Systems, 2007.

[5] I. Cox, J. Kilian, F. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1673–1687, 1997.

[6] F. Hartung, "Spread spectrum watermarking: Malicious attacks and counterattacks," *Proc. SPIE*, 1999.

[7] B. Chen and G. Wornell, "Quantization index modulation methods for digital watermarking and information embedding of multimedia," *The Journal of VLSI Signal Processing*, 2001.

[8] P. Moulin and A. K. Goteti, "Block qim watermarking games," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 3, pp. 293–310, Sep 2006.

[9] L. Perez-Freire, "Spread-spectrum vs. quantization-based data hiding: misconceptions and implications," *Proc. SPIE*, vol. 2, 2005.

[10] T. Kalker, "Considerations on watermarking security," in *Multimedia Signal Processing, 2001 IEEE Fourth Workshop on*. IEEE, 2001, pp. 201–206.

[11] T. Furon, J. Oostven, and J. Bruggen, "Security analysis," in *Deliverable D.5.5, CERTIMARK IST European Project*, 2002.

[12] F. Cayre, C. Fontaine, and T. Furon, "Watermarking security: theory and practice," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3976–3987, Oct 2005.

[13] P. Comesaña, L. Pérez-Freire, and F. Pérez-González, "Fundamentals of data hiding security and their application to spread-spectrum analysis," *Information Hiding*, 2005.

[14] L. Pérez-Freire and F. Pérez-González, "Spread-spectrum watermarking security," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 1, pp. 2–24, Mar 2009.

[15] P. Bas and A. Westfeld, "Two key estimation techniques for the broken arrows watermarking scheme," in *Proceedings of the 11th ACM Workshop on Multimedia and Security*. ACM Press, 2009, p. 1.

[16] F. Xie, T. Furon, and C. Fontaine, "Towards Robust and Secure Watermarking," in *ACM Multimedia and Security*, 2010.

[17] A. Kerckhoffs, "La cryptographie militaire," *J. Sci. Milit.*, vol. IX, pp. 5–38 and 161–191, Jan 1883.

[18] I. J. Cox, G. Doerr, and T. Furon, "Watermarking is not cryptography," *Lecture Notes in Computer Science*, 2006.

[19] P. Bas, T. Furon, and F. Cayre, "Practical Key Length of Watermarking Systems," *IEEE ICASSP*, 2012.

[20] M. Barni, F. Bartolini, and T. Furon, "A general framework for robust watermarking security," *Signal Process.*, vol. 83, pp. 2069–2084, 2003.

[21] F. Cayre and P. Bas, "Kerckhoffs-based embedding security classes for woa data hiding," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, pp. 1–15, 2008.

[22] A. D. Ker, *Perturbation Hiding and the Batch Steganography Problem*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2008, vol. 5284, ch. chapter 4, pp. 45–59.

[23] C. Cachin, "An information-theoretic model for steganography," *Information Hiding*, vol. 1525, Nov. 1998.

[24] I. S. Moskowitz, G. E. Longdon, and L. Chang, "A new paradigm hidden in steganography," in *Proceedings of the 2000 Workshop on New Security Paradigms*. ACM Press, 2000, pp. 41–50.

[25] B. Mathon, P. Bas, F. Cayre, and B. Macq, "Comparison of secure spread-spectrum modulations applied to still image watermarking," *Annals of Telecommunications - Annales Des Télécommunications*, Jul. 2009.

[26] L. Pérez-Freire and F. Pérez-González, "Security of lattice-based data hiding against the watermarked-only attack," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 4, pp. 593–610, Dec 2008.

[27] B. Mathon, P. Bas, and F. Cayre, "Practical performance analysis of secure modulations for woa spread-spectrum based image watermarking," in *Proceedings of the 9th Workshop on Multimedia & Security*. ACM Press, 2007, p. 237.

[28] P. Bas and F. Cayre, *Natural Watermarking: A Secure Spread Spectrum Technique for WOA*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2007, vol. 4437, ch. chapter 1, pp. 1–14.

[29] ——, "Achieving subspace or key security for woa using natural or circular watermarking," in *Proceedings of the 8th workshop on Multimedia and security*. ACM, 2006, pp. 80–88.

[30] C. Villani, *Optimal transport: old and new*. Springer Berlin Heidelberg, 2009.

[31] B. Mathon, P. Bas, F. Cayre, and B. Macq, "Optimization of natural watermarking using transportation theory," in *Proceedings of the 11th ACM Workshop on Multimedia and Security*. ACM Press, 2009, p. 33.

[32] P. Bas, "Soft-SCS: improving the security and robustness of the Scalar-Costa-Scheme by optimal distribution matching," *Information Hiding*, 2011.

[33] J. Cao and J. Huang, "Controllable secure watermarking technique for tradeoff between robustness and security," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 821–826, Apr 2012.

[34] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[35] B. Mathon, P. Bas, F. Cayre, and F. Pérez-González, *Distortion Optimization of Model-Based Secure Embedding Schemes for Data-Hiding*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2008, vol. 5284, ch. chapter 23, pp. 325–340.

[36] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.

[37] J. H. Conway, R. H. Harding, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in grassmannian spaces," *Exp. Math.*, vol. 5, no. 2, pp. 139–159, 1996.

[38] M. Matsumoto and T. Nishimura, "Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator," *ACM Trans. Modeling Computer Simulation*, vol. 8, no. 1, pp. 3–30, 1998.

[39] G. Marsaglia and W. W. Tsang, "The ziggurat method for generating random variables," *J. Stat. Softw.*, vol. 5, no. 8, pp. 1–7, 2000.

[40] G. Box and M. Müller, "A note on the generation of random normal deviates," *Ann. Math. Statist.*, vol. 29, no. 2, pp. 610–611, 1958.

[41] E. R. Berlekamp, *Nonbinary BCH decoding*. University of North Carolina. Dept. of Statistics, 1966.

[42] J. L. Massey, "Shift-register synthesis and bch decoding," *IEEE Trans. Inf. Theory*, vol. IT-15, no. 1, pp. 122–127, Jan 1969.

[43] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Res. Logistic Q.*, vol. 2, pp. 83–97, 1955.

[44] G. Monge, "Mémoire sur la théorie des déblais et des remblais," *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, pp. 666–704, 1781.

[45] L. Kantorovitch, "On the translocation of masses," *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, vol. 37, pp. 199–201, 1942.

[46] L. Kantorovich, "On a problem of monge," *Uspekhi Mat. Nauk.*, vol. 3, pp. 225–226, 1948.

[47] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems, Volume I: Theory*. Springer-Verlag, 1998.

[48] M. Knott and C. S. Smith, "On the optimal mapping of distributions," *Journal of Optimization Theory and Applications*, vol. 43, no. 1, pp. 39–49, May 1984.

[49] N. Wiener, "Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications," *Journal of the American Statistical Association*, 1949.

[50] A. Cohen, I. Daubechies, and J.-C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Communications on pure and applied mathematics*, vol. 45, no. 5, pp. 485–560, 1992.

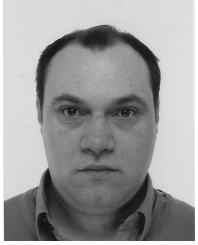
[51] B. E. Usevitch, "Optimal bit allocation for biorthogonal wavelet coding," in *Data Compression Conference, 1996. DCC '96. Proceedings*, Mar 1996, pp. 387–395.

[52] ——, "A tutorial on modern lossy wavelet image compression : Foundations of jpeg 2000," *IEEE Signal Process. Mag.*, vol. 18, no. 9, pp. 22–35, Sep 2001.

[53] P. Bas and T. Furon, "Break Our Watermarking System 2nd edition," http://bows2.ec-lille.fr.

[54] A. Piva, M. Barni, F. Bartolini, and V. Cappellini, "DCT-Based Watermark Recovering Without Resorting to the Uncorrupted Original Image," in *IEEE Signal Processing Society 1997 International Conference on Image Processing (ICIP'97)*, Santa Barbara, California, 1997.

Benjamin Mathon received the M.S. degree in cryptology, security and information coding from the Université Joseph Fourier, Grenoble, France, in 2007. In 2011, he received the Ph.D. degree from both the Université catholique de Louvain, Louvain-la-Neuve, Belgium, and the Institut polytechnique de Grenoble, France in 2011. He was a post-doctoral fellow at Institut polytechnique de Grenoble and Université de Genève, Switzerland from 2011 to 2012 and post-doctoral fellow at Institut National de Recherche en Informatique et Automatique, Rennes, France, from 2012 to 2013. Since September 2013, Benjamin Mathon works at Université de Lille 1 as assistant professor. His main interests include image, information theory, watermarking, traitor tracing and multimedia security.



François Cayre got his Ph.D. from Telecom ParisTech, Paris, France and Université catholique de Louvain, Louvain-la-Neuve, Belgium in 2003. He was a post-doctoral fellow at Institut National de Recherche en Informatique et Automatique, Rennes, France until 2005 when he joined Grenoble-INP as an assistant professor. His interests include watermarking security and multimedia security at large.

Patrick Bas received the Electrical Engineering degree from the Institut National Polytechnique de Grenoble, France, in 1997 and the Ph.D. degree in Signal and Image processing from Institut National Polytechnique de Grenoble, France, in 2000. From 1997 to 2000, he was a member of the Laboratoire des Images et des Signaux de Grenoble (LIS), France where he worked on still image watermarking. During his post-doctoral activities, he was a Member of the Communications and Remote Sensing Laboratory of the Faculty of Engineering at the Université Catholique de Louvain, Belgium. His research interests include synchronisation and security evaluation in watermarking, and steganalysis. From 2004 to 2008, Patrick Bas is co-coordinator of the virtual lab 1 on "watermarking and theory" within the Ecrypt European NoE. In 2007, Patrick Bas has co-organised the 9th International Workshop on Information Hiding (IH07), the 3rd Wavila Challenge, the 2nd Edition of the Bows-2 contest on Watermarking and the first edition of the BOSS contest on steganalysis. From 2005 to 2008, Patrick Bas was detached from Gipsa-Lab to work as a visiting researcher at the Computer and Information Science Laboratory in Aalto University (Finland). From 2001 to 2009, Patrick Bas worked as a CNRS researcher at Gipsa-Lab, and since 2010 he works at LAGIS, Lille, France.

Benoît Macq is currently Professor at Université catholique de Louvain, in the Telecommunication Laboratory and Vice-Rector of the University. He completed his military service in 1984-1985 at the Royal Military School of Belgium where he worked on Laser interferometer measurements. He worked on networks planning in 1985 for the Tractebel company, Brussels. He did his doctoral thesis on perceptual coding for digital TV at UCL. He was researcher at Philips Research in 1990 and 1991. He has been senior researcher of the Belgian NSF. Benoit Macq has been visiting scientist at École Polytechnique Fédérale de Lausanne and at the Massachusetts Institute of Technology, Boston. He has been Scientific Visitor at MIT and Visiting Professor at the École Nationale Supérieure des Télécommunications, ENST-Paris, France as well as at the Université de Nice Sophia-Antipolis, France. Benoit Macq is teaching and doing his research work in image processing for visual communications. He is closely collaborating with the MIT, Boston, USA, the Harvard Medical School, Boston, USA, and with the NARA Institute, Japan. His main research interests are image compression, image watermarking, image analysis for medical and immersive communications.